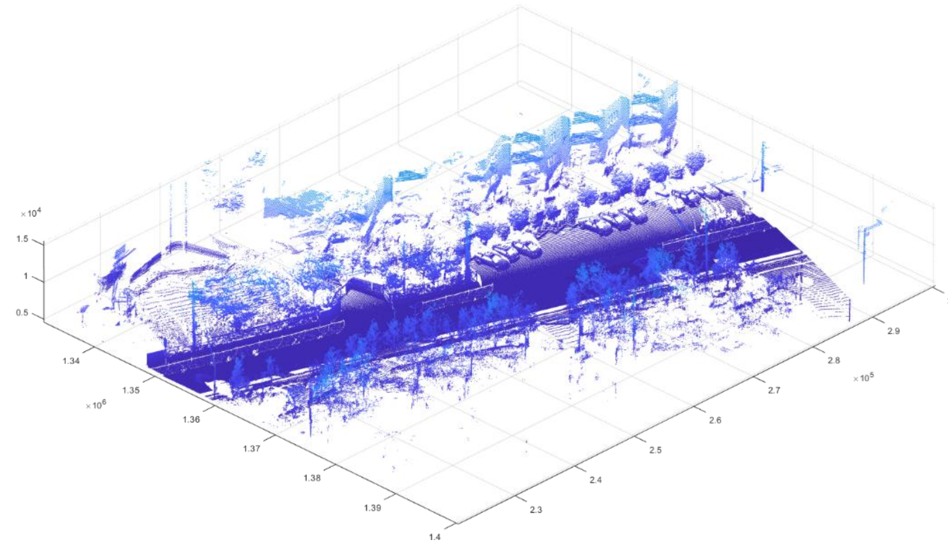
Point Cloud Compression & Communication

Zhu Li
Director, UMKC NSF Center for Big Learning
Dept of Computer Science & Electrical Engineering
University of Missouri, Kansas City
Email: zhu.li@ieee.org, lizhu@umkc.edu
Web: http://l.web.umkc.edu/lizhu



Outline

- Short Self Intro
- Research Motivation and Highlights
- Scalable Point Cloud Geometry Compression
- Video Projection Based Point Cloud Compression
- Post-Processing: Point Cloud Scaling
- Summary
- URL: http://l.web.umkc.edu/lizhu/docs/pccc.pdf

Short Bio:

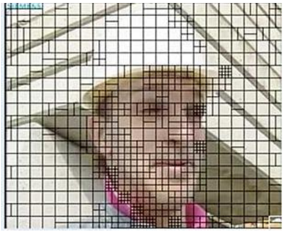
NSF I/UCRC Center *for* Big Learning
Creating Intelligence

Research Interests:

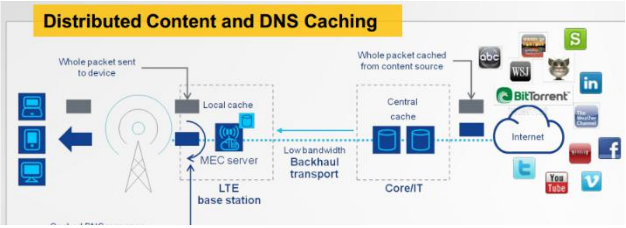
- Immersive Media Communication: light field, point cloud and 360 video capture, coding and low latency communication.
- Data & Image Compression: video, medical volumetric data, DNA sequence, and graph signal compression with deep learning
- Remote Sensing & Vision: vision problem under low resolution, blur, and dark conditions, hyperspectral imaging, sensor fusion
- Edge Computing & Federated Learning: gradient compression, light weight inference engine, retina features, fast edge cache for video CDN

signal processing and learning

image understanding



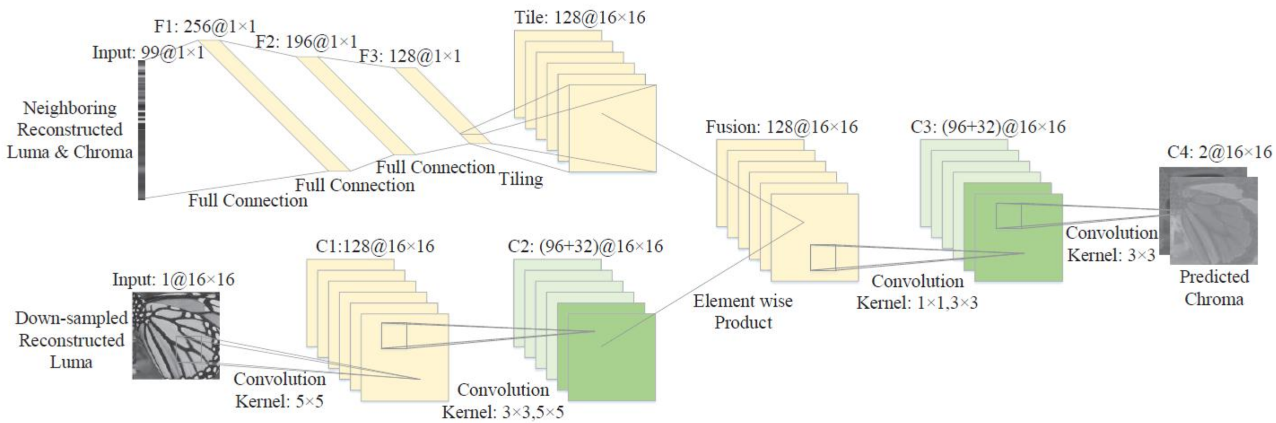
visual communication



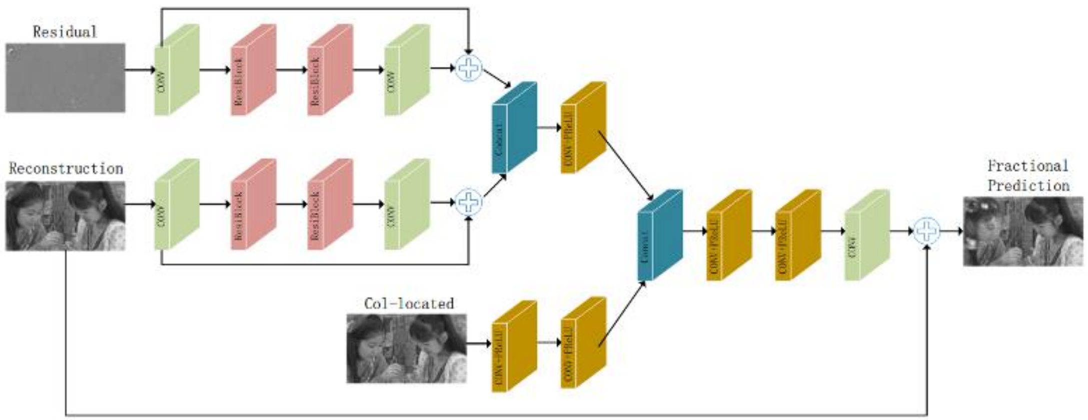
mobile edge computing & communication

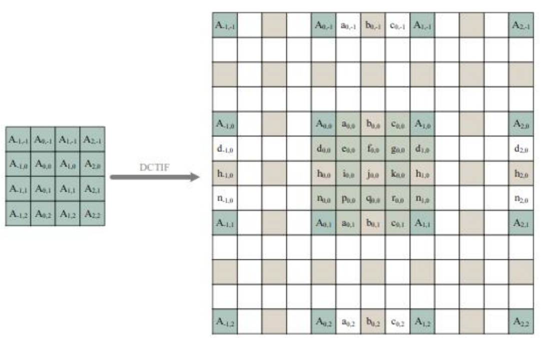
Data & Image Compression Highlights (NSF/IUCRC)

- “Neural Network Based Cross-Channel Intra Prediction”, *ACM Trans on Multimedia Computing Communication and Applications* (TOMM), 2021.



- “Compression Priors Assisted Convolutional Neural Network for Fractional Interpolation”, *IEEE Trans on Circuits and Systems for Video Tech*. (T-CSVT), 2020





Edge Media Computing & Federated Learning

- "Referenceless Rate-Distortion Modeling with Learning from Bitstream and Pixel Features", *ACM Multimedia* (MM), Seattle, 2020.

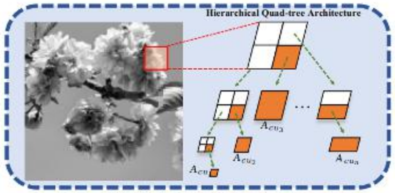


Fig 4. Segmentation Mappings

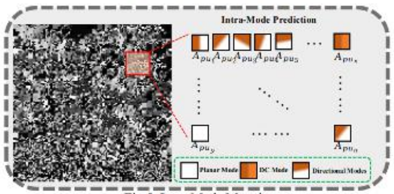
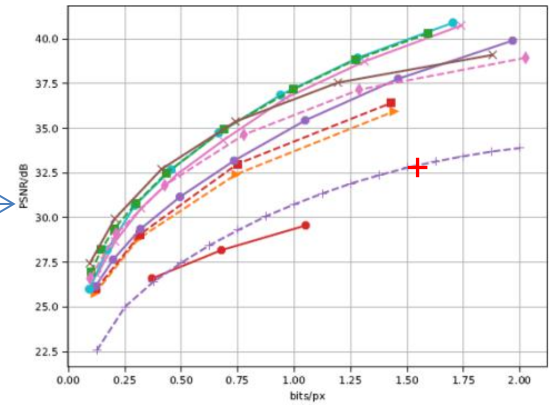


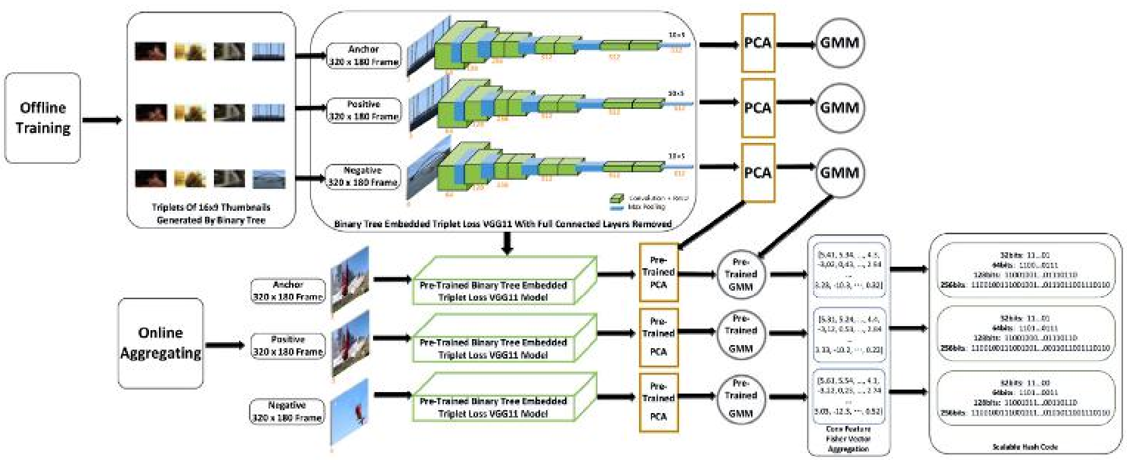
Fig 5. Intra-Mode Mappings

learn from
one encoding

ref-less R-D modeling

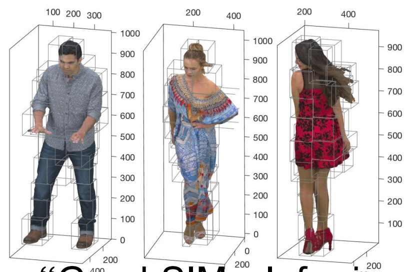


- “Scalable Hash From Triplet Loss Feature Aggregation for Video De-Duplication”, *Journal of Visual Communication & Image Representation* (JVCIR), 2020.



0-FPR deduplication at <10ms latency for very large repository

Immersive Media Coding & Communication (NSF/IUCRC)




- “GraphSIM - Inferring Point Cloud Quality via Graph Similarity”, *IEEE Trans on Pattern Analysis & Machine Intelligence* (T-PAMI), 2021.
- "Efficient Projected Frame Padding for Video-based Point Cloud Compression", *IEEE Trans on Multimedia*(T-MM), 2020.
- "Rate Control for Video-based Point Cloud Compression", *IEEE Transactions on Image Processing* (T-IP), 2020.
- "λ-domain Perceptual Rate Control for 360-degree Video Compression", *IEEE Journal of Selected Topics in Signal Processing* (JSTSP), 2020.
- "Advanced 3D Motion Prediction for Video Based Dynamci Point Cloud Compression", *IEEE Trans on Image Processing*(T-IP), 2019.
- "Quadtree-based Coding Framework for High Density Camera Array based Light Field Image", *IEEE Trans on Circuits and Systems for Video Tech*(T-CSVT), 2019.
- "Advanced Spherical Motion Model and Local Padding for 360 Video Compression", *IEEE Trans on Image Processing* (T-IP) vol. 28, no. 5, pp. 2342-2356, May 2019.
- “Scalable Point Cloud Geometry Coding with Binary Tree Embedded Quadtree”, *IEEE Int'l Conf. on Multimedia & Expo* (ICME) ,San Diego, USA, 2018.
- “Pseudo sequence based 2-D hierarchical coding structure for light-field image compression”, *IEEE Journal of Selected Topics in Signal Processing* (JSTSP), Special Issue on Light Field, 2017.

Outline

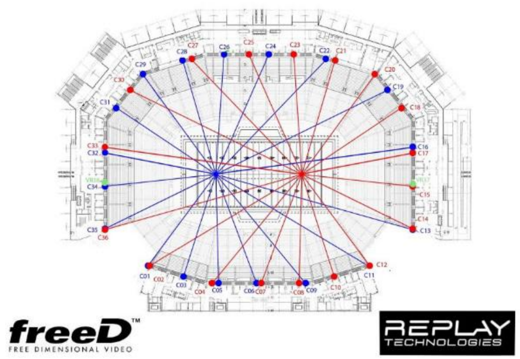
- Short Self Intro
- Research Motivation and Highlights
- Scalable Point Cloud Geometry Compression
- Video Projection Based Point Cloud Compression
- Post-Processing: Point Cloud Scaling
- Summary

What is Point Cloud

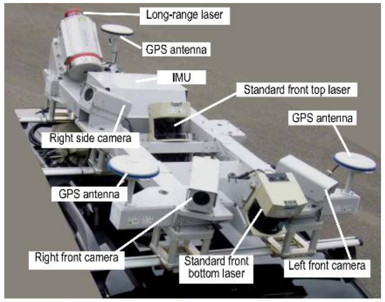
- A collection of Un-ordered points with
  - Geometry: expressed as [x, y, z]
  - Color Attributes: [r g b], or [y u v]
  - Additional info: normal, timestamp, …etc.
- Key difference from mesh: no order or local topology info

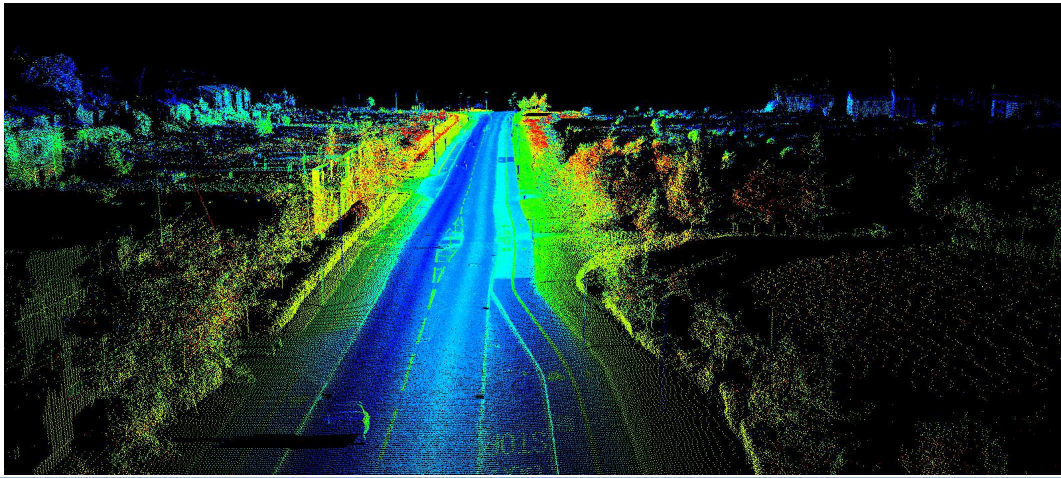
Point Cloud Capture

- Passive: Camera array stereo depth senso



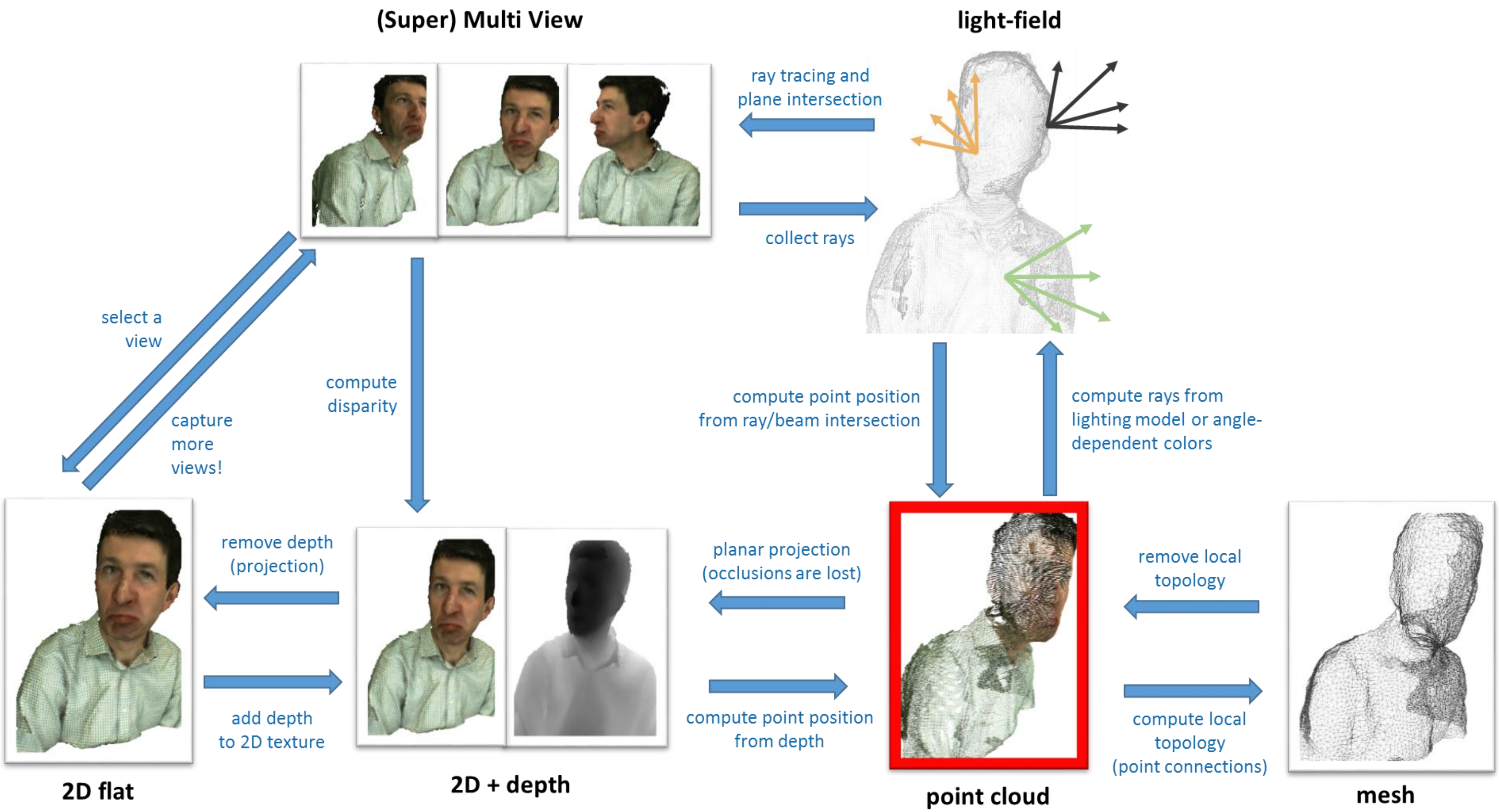
- Active: LiDAR, mmWave, TOF sensors





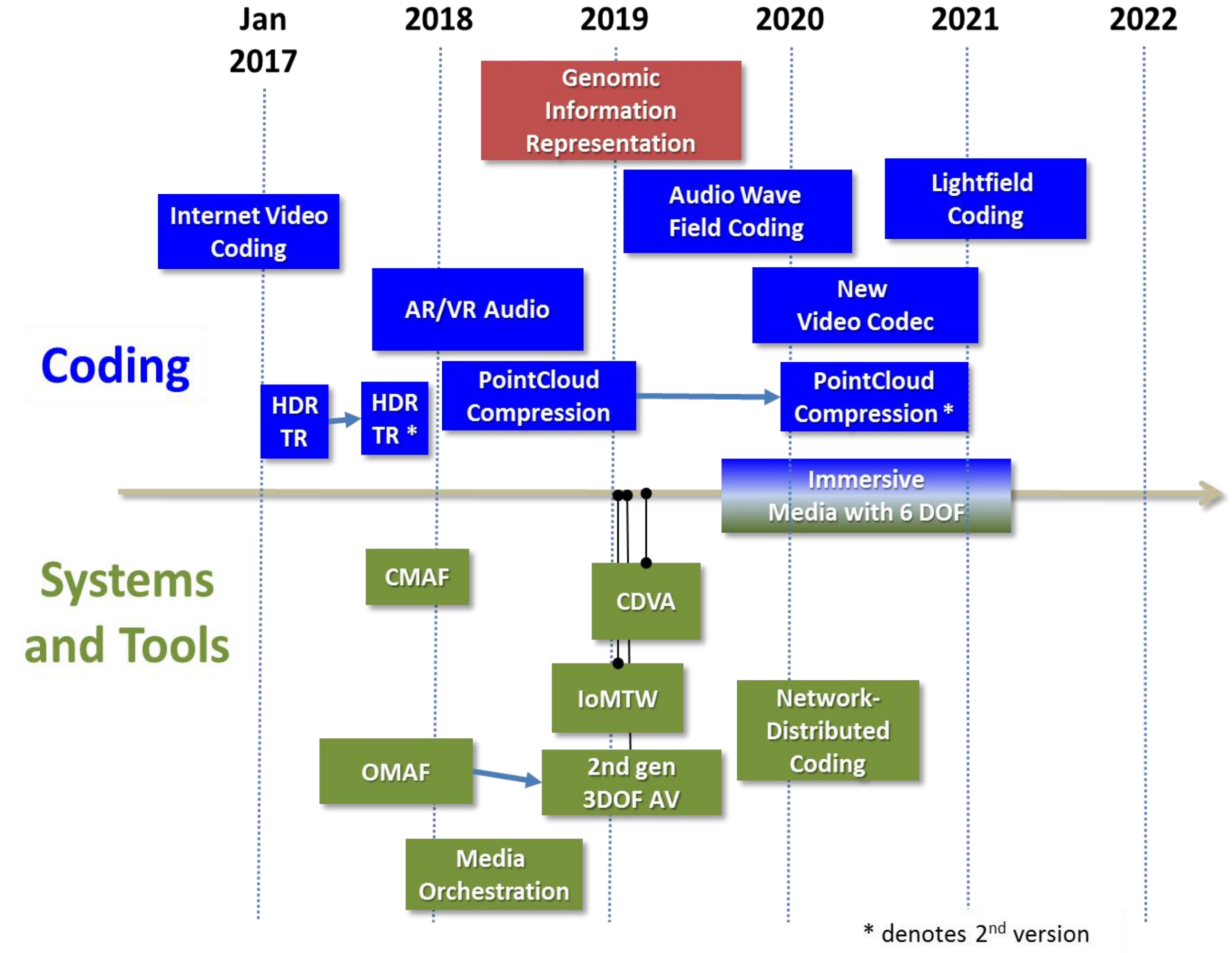
Point Cloud Inter-Operability with Other Formats

- Provide true 6-DoF Content capacity



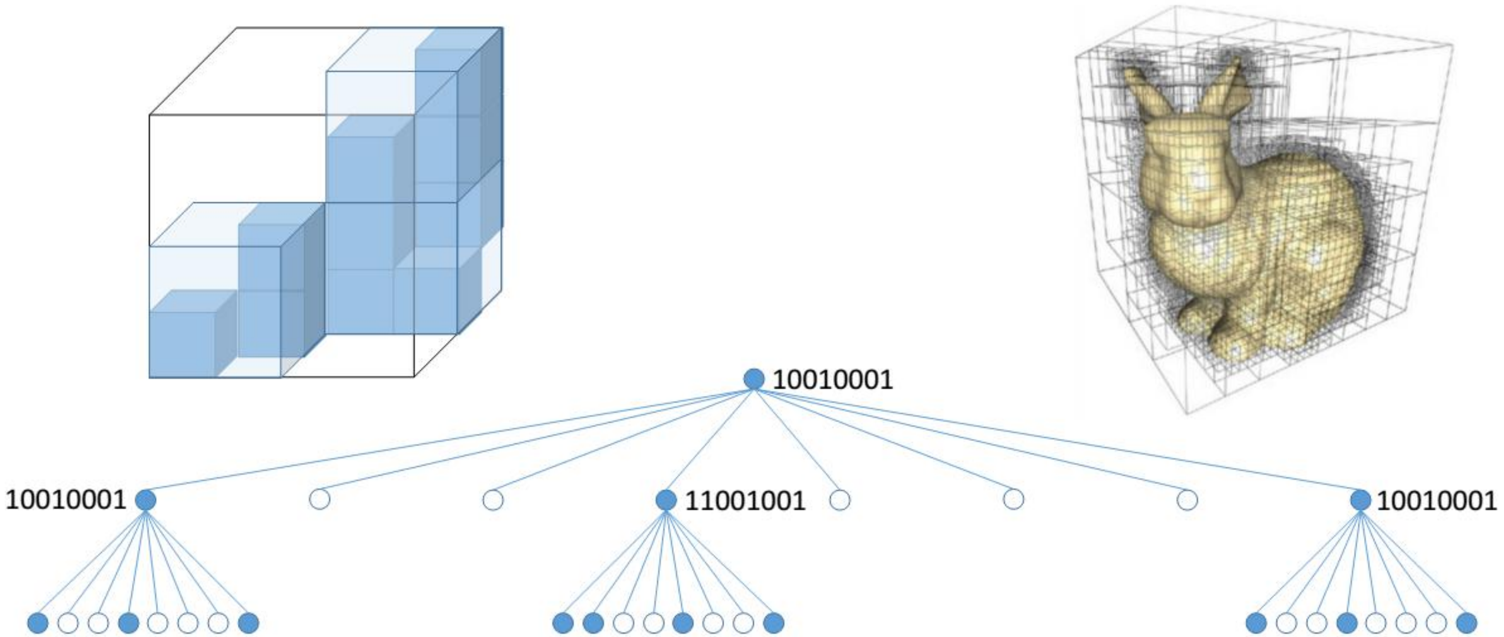
PCC in MPEG

- Part of the MPEG-Immersive grand vision



Octree Based Point Cloud Compression

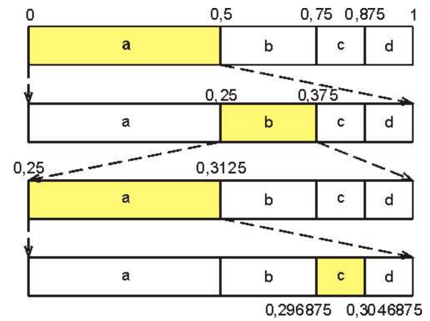
- Octree is a space partition solution
  - Iteratively partition the space into sub-blocks.
  - Encoding: 0 if empty, 1 if contains data points
  - Level of the tree controls the quantization error

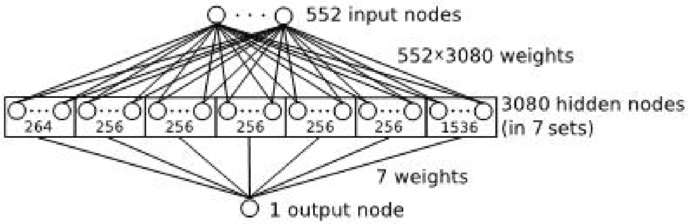


Credit: Phil Chou, PacketVideo 2016

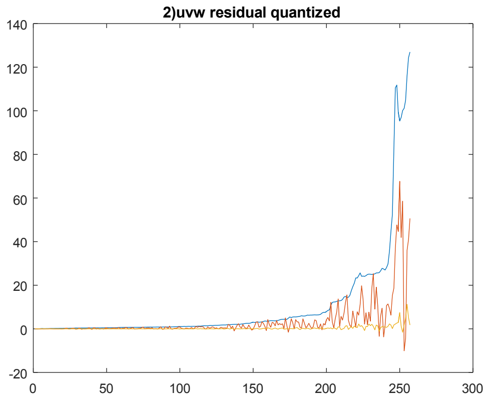
Lossless Compression of the Octree with Neural Network driven CABAC

- Tree Structure:
  - DFS scanning of the Octree node byte to have a byte stream
  - Compression of the byte stream via Arithmetic Coding, or shallow neural network PAQ coding





- Residual Coding:
  - Range coding: coding the residual against a ref point (eg., centroids of octree leaf node centroids)
  - Plane/Surface approximation coding: compute the projection distances to a surface, surface can be polynomial or planar.

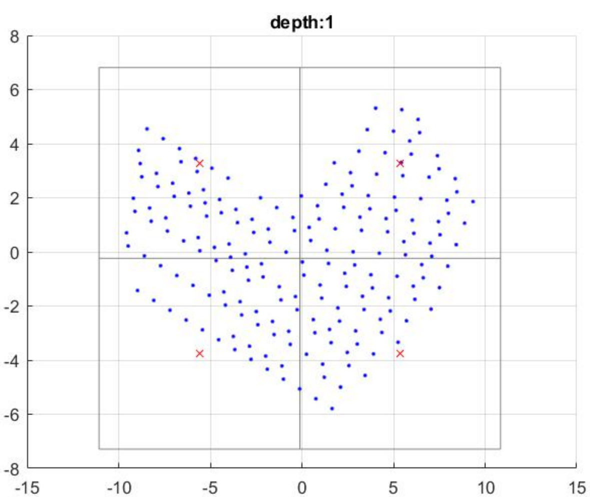


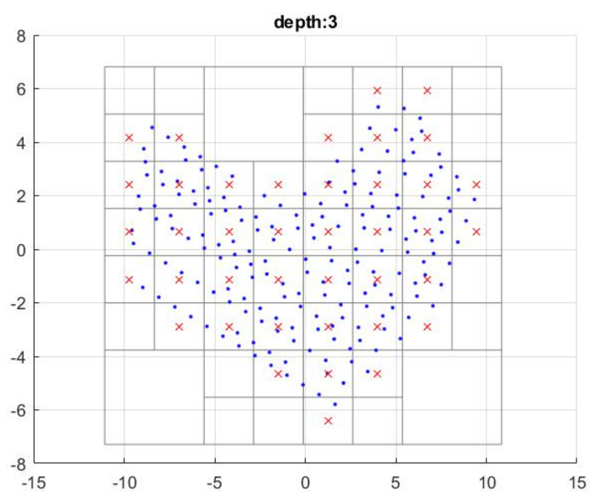
Scalable Point Cloud Geometry Coding

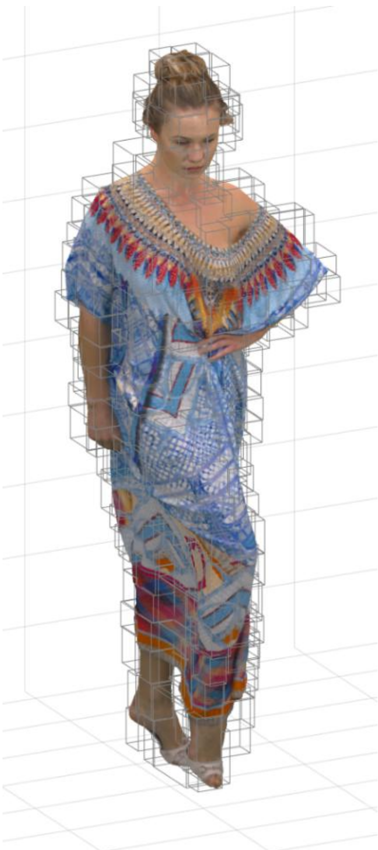
- Binary Tree embedded Quadtree (BTQT) coding (adopted in MPEG gPCC):
  - Binary tree partition to have lossy geometry approximation
  - Refine each leaf node with Quadtree/Octree to offer scalable details upto near lossless

Scalable Geometry Coding with BTQT

- Construct Binary Tree of Point Cloud
  - $R_1 = (2^L-1)*(2+K) + 6*K$, cost of signalling for resolution *K* bits and binary tree depth *L*
- *Intra-Coding* i.e. either *Quadtree* (flat surface) or *Octree*(not flat)
  - QT case overhead: R_2 = 3*p + 3*q bits, for singalling norma at p bits and point at q bits. q < K *proportional* to the leaf node size.







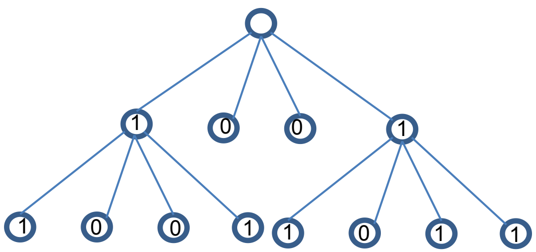
Quadtree/Octree Mode Decision

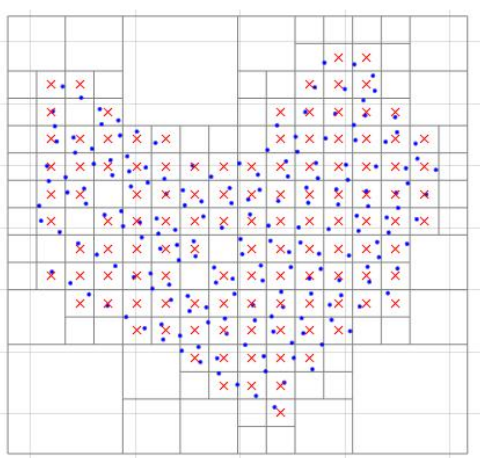
Quadtree | Octree

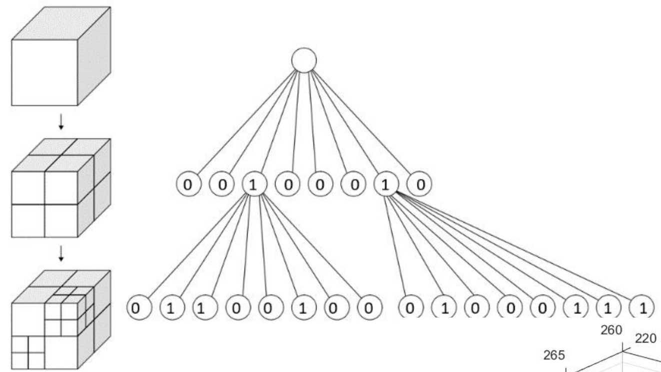
Flatness Criterian:

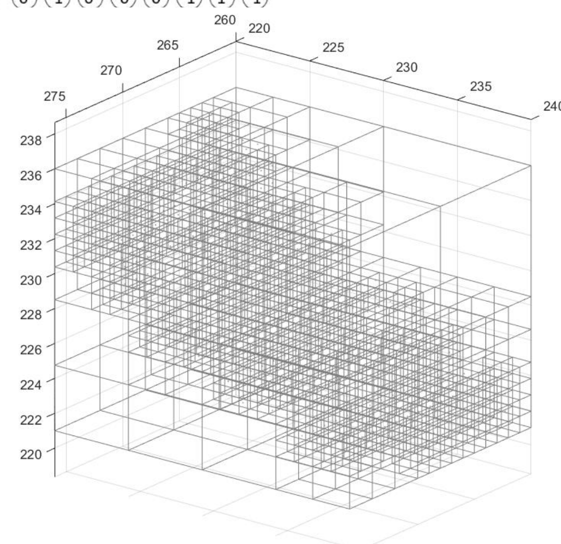
$$\theta = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}$$

$$\lambda = eig(cov(X))$$



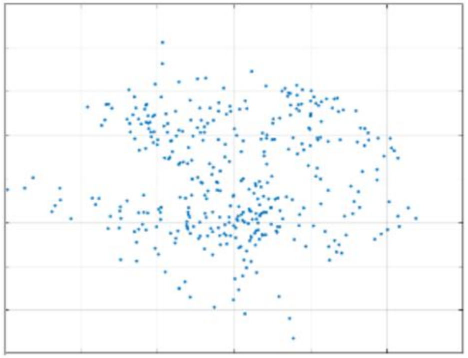


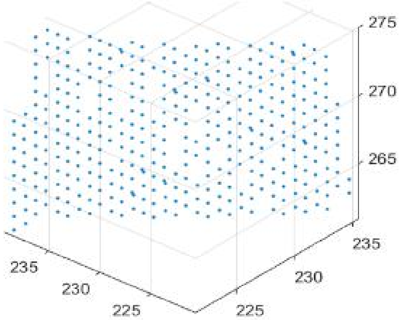




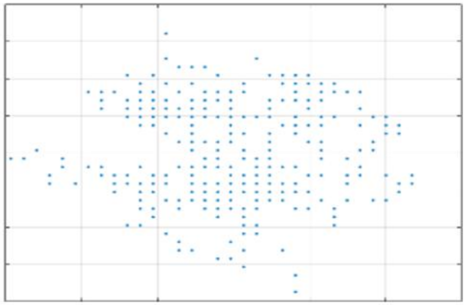
Scalable Coding with Quadtree/Octree

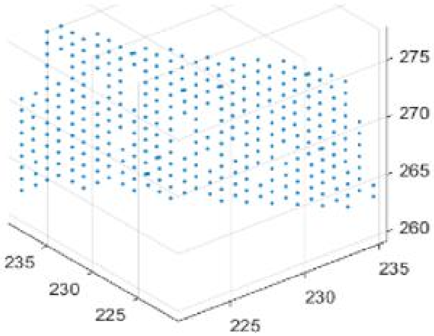
Quadtree | **Octree**



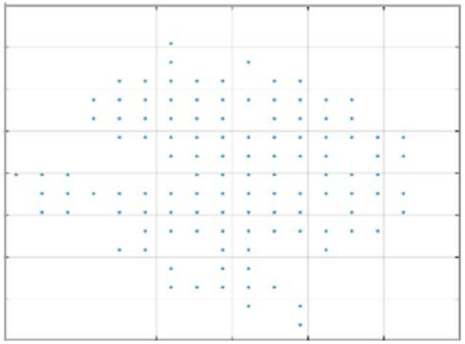


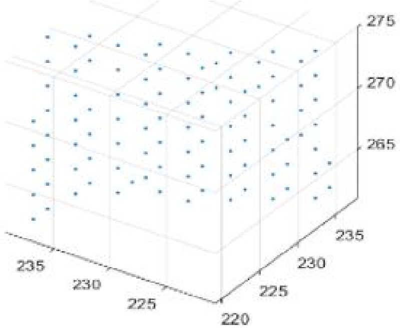
L = L1





L = L2 < L1

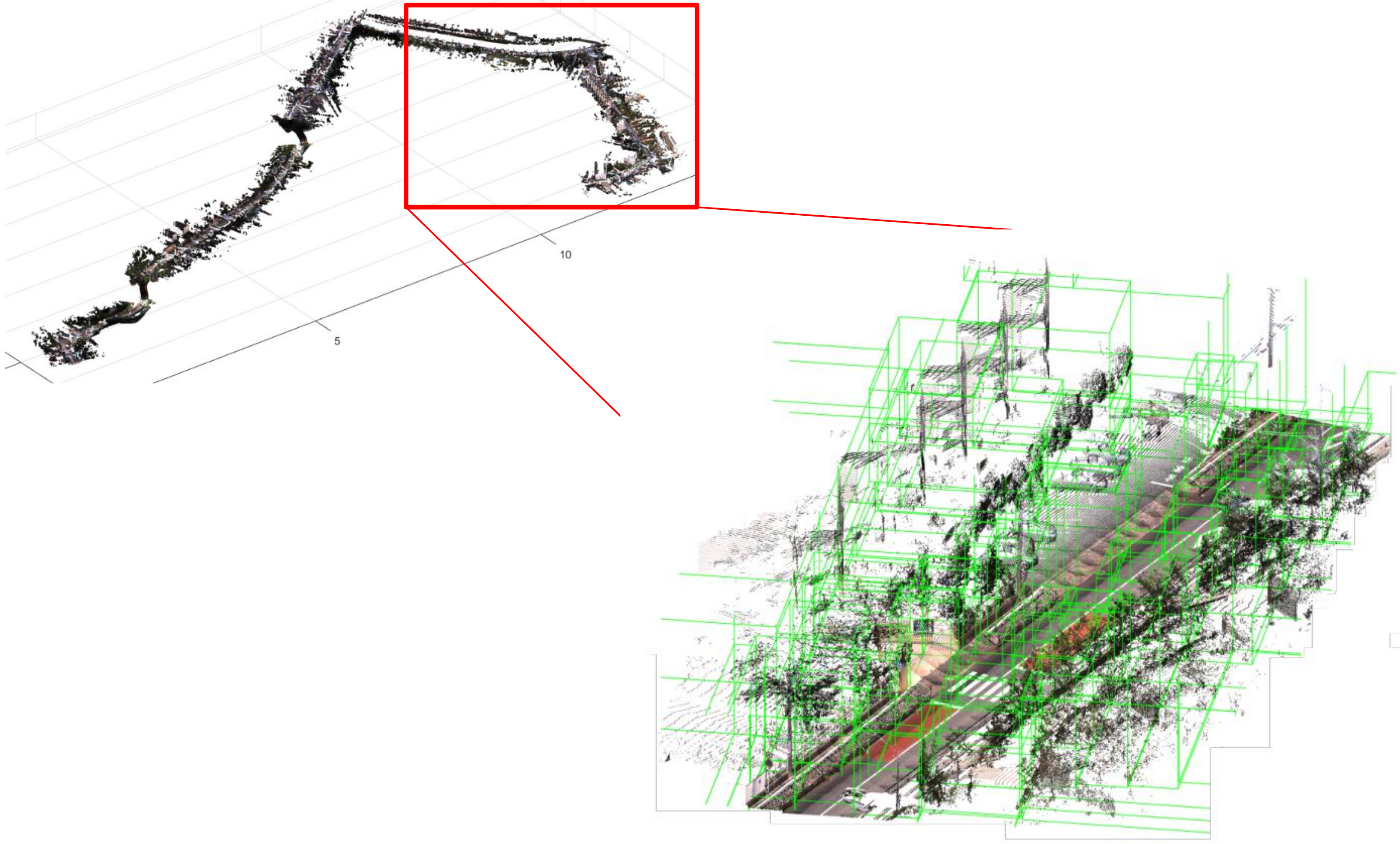




L = L3 < L2

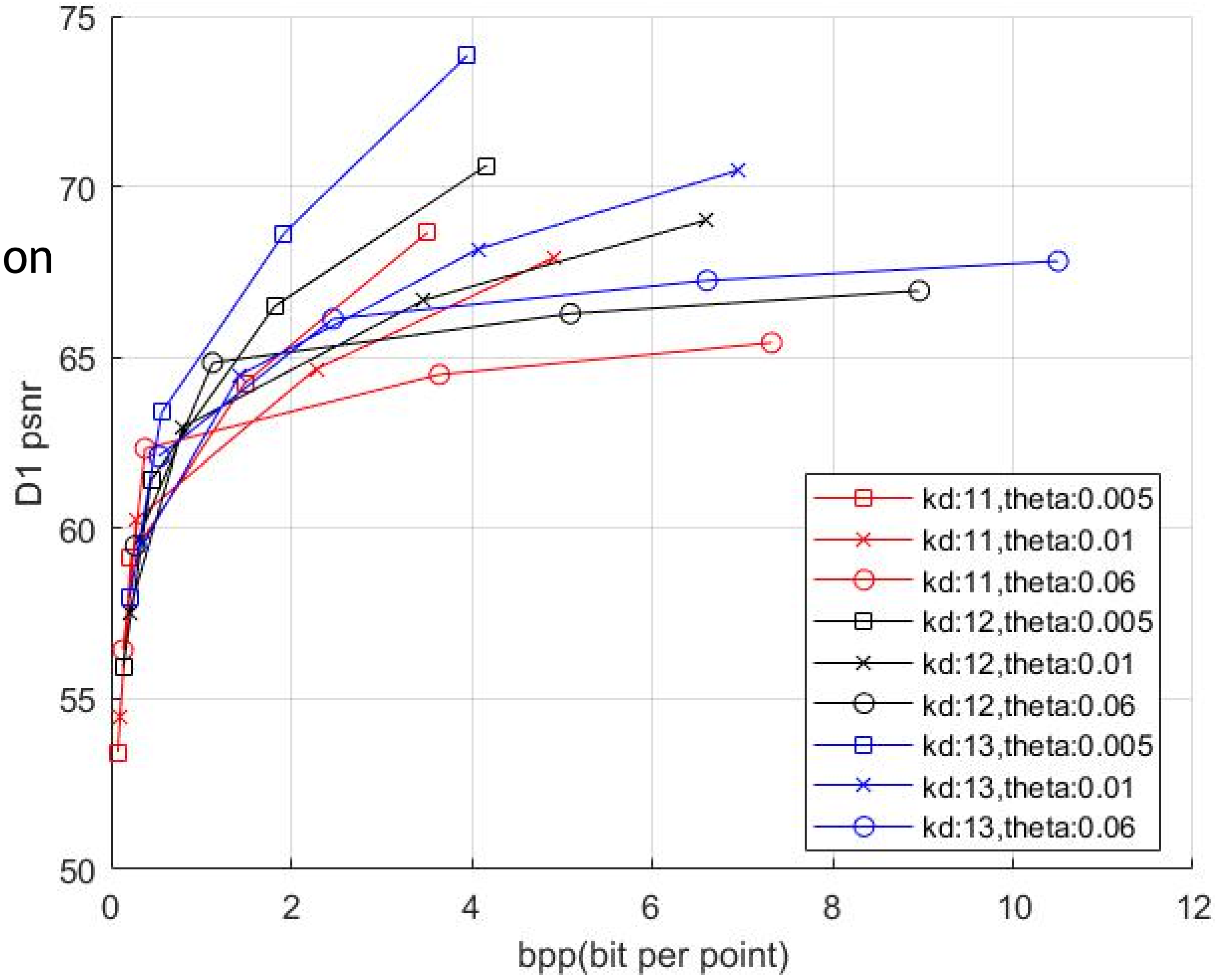
Point cloud Visualization

- *citytunnel* dataset (MERL) – 1.5 km long section of a



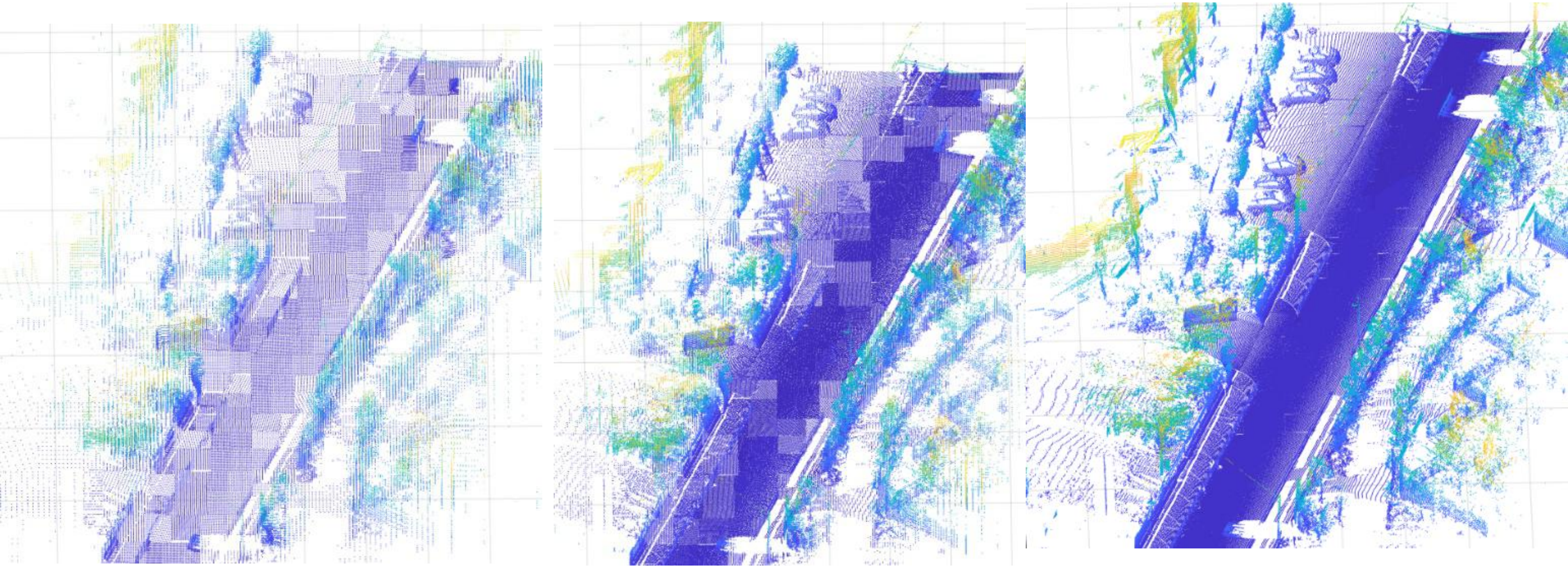
Result: Category 1 Geometry Coding Efficiency

near lossless
at 30x compression



Reconstructed-Zoomed

- Various reconstruction accuracy:

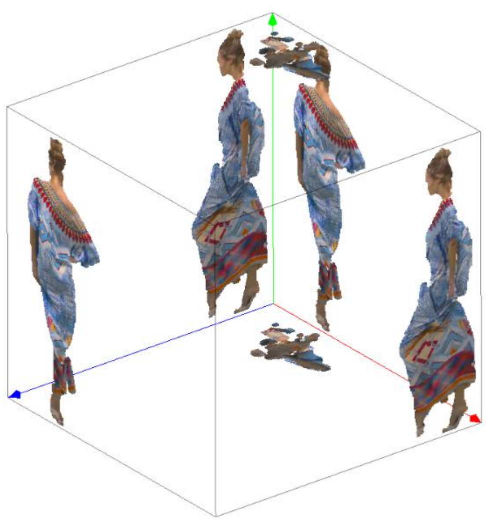


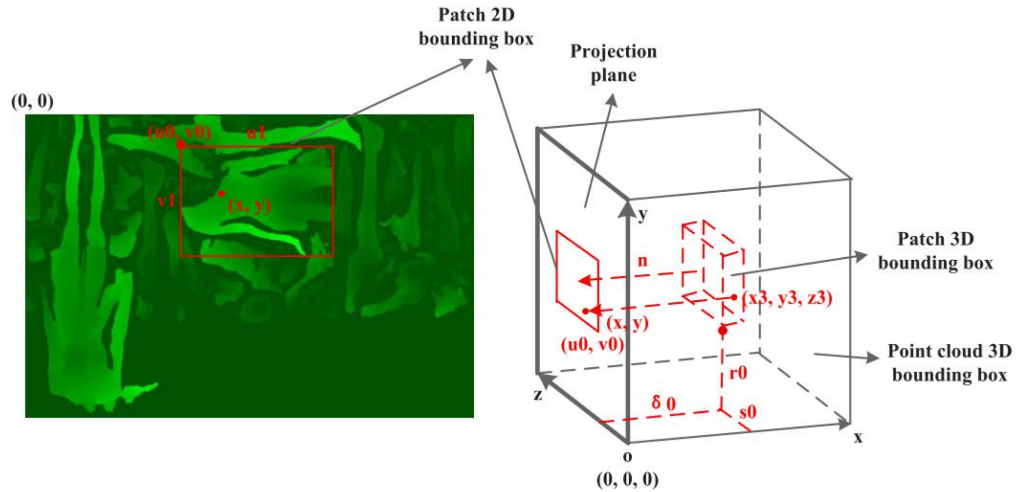
Outline

- Short Self Intro
- Research Motivation and Highlights
- Scalable Point Cloud Geometry Compression
- Video Projection Based Point Cloud Compression
- Post-Processing: Point Cloud Scaling
- Summary

Video-based point cloud compression

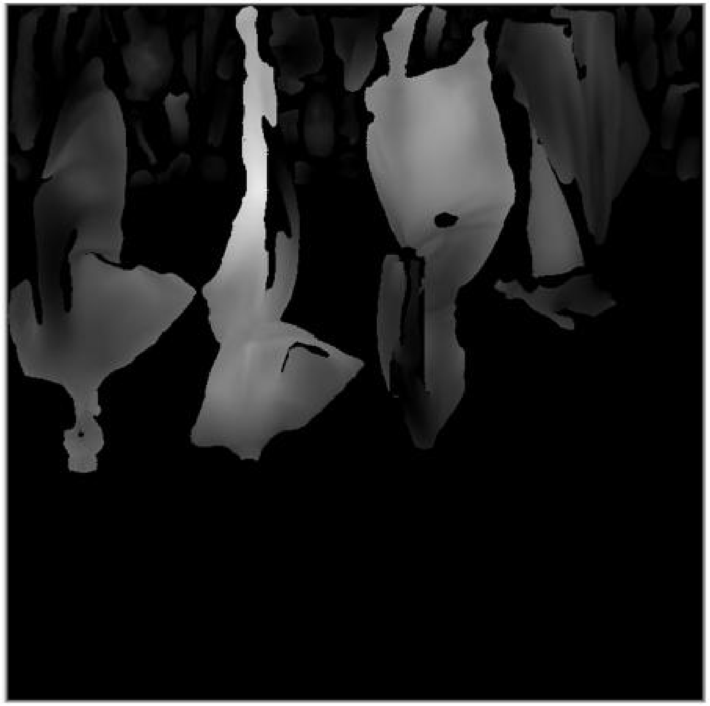
- Basic steps
  - Normal-based projection, frame packing, and frame padding
- Normal-based projection
  - Organize the points with similar normal into a patch
  - Project each patch to the 3D point cloud bounding box





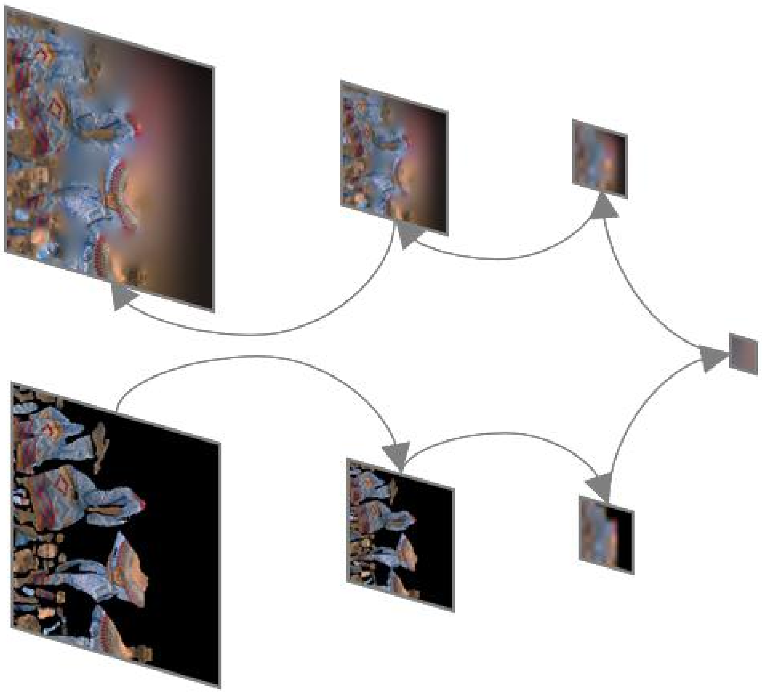
Video-based point cloud compression

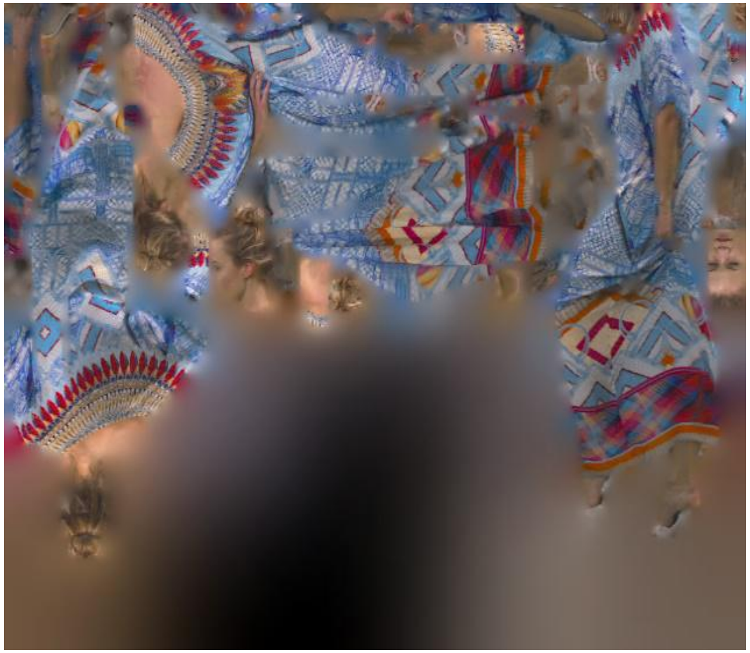
- Frame packing: pack the patches into frames
  - Exhaustive search empty space for the current patch
  - Patch rotation is supported
  - Introduced a lot of sharp edges



VPCC - Texture Padding

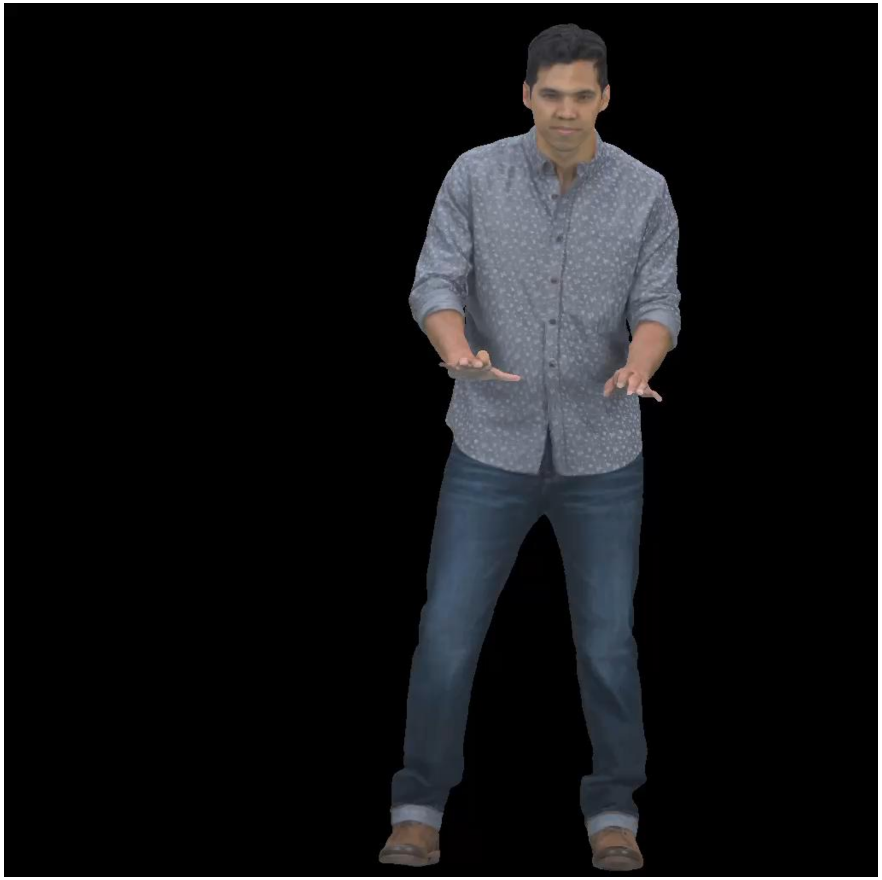
- Texture padding: a number of methods are proposed to minimize the bitrate of the unoccupied pixels
- Using push-pull algorithm as an example, like dilation



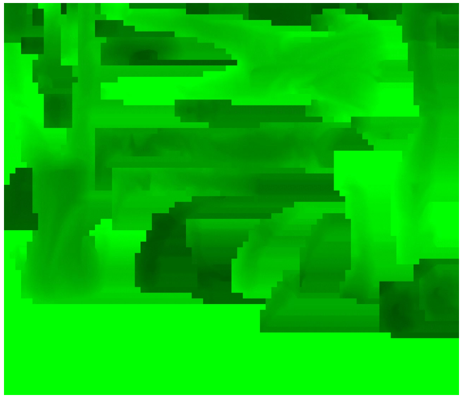


Video-based point cloud compression

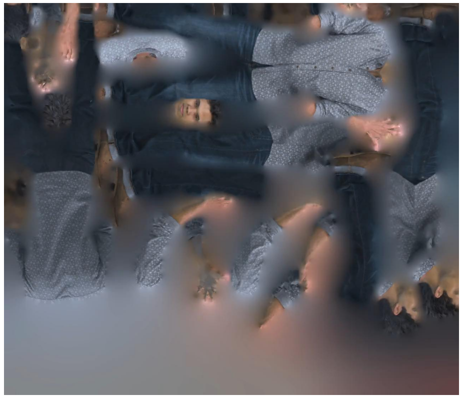
- Basic idea: project a point cloud to a 2-D video for an efficient compression (demo)



Geometry



Attribute



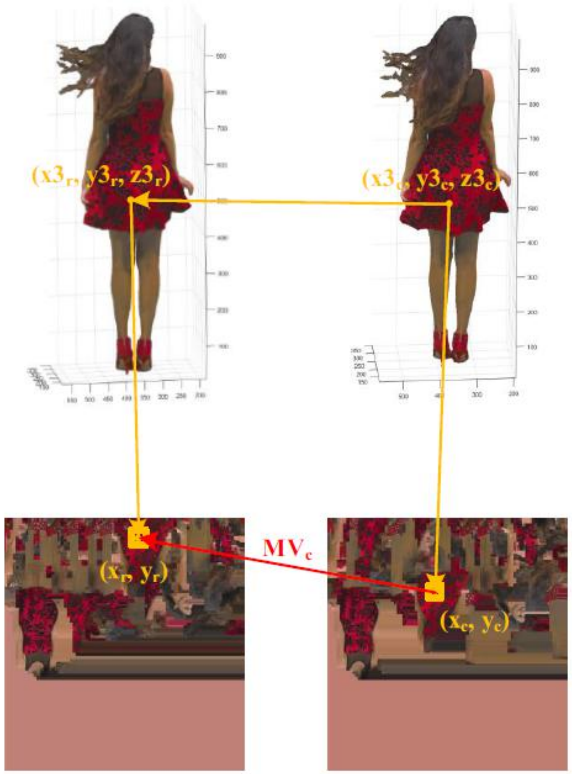
VPCC Motion Model

- The corresponding patches may be put in totally different positions in various frames (Green squares)
  - The current video codec may be unable to find a good motion vector for each block in this case
  - The geometry is encoded before the attribute, we can use the geometry to derive a better motion vector for attribute


General 3D to 2D motion model

- Given the 3D motion and the 3D to 2D correspondence, we can derive the 2D motion
  - g(), f(): 3D to 2D projection in reference and current frames

$$\boldsymbol{MV_c} = \boldsymbol{g}(x3_r, y3_r, z3_r) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

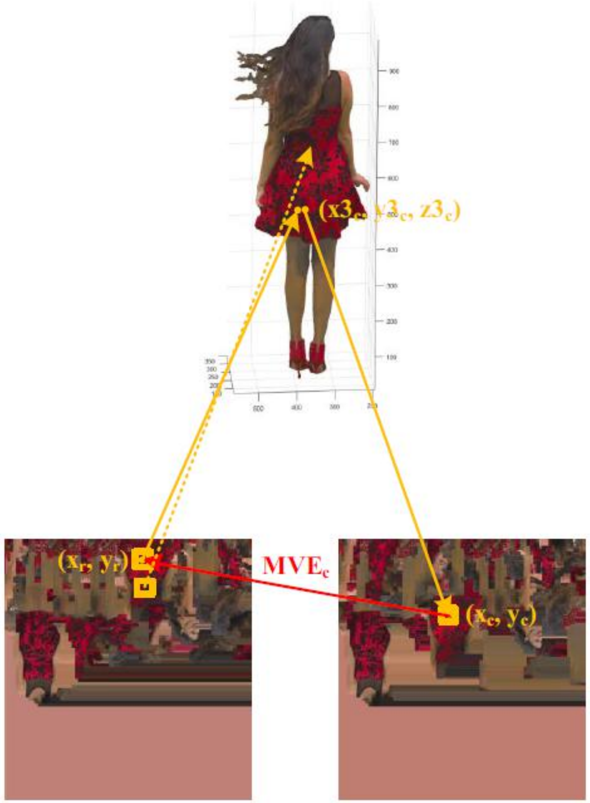


Geometry-based motion prediction

- In the V-PCC, we know the 3D-to-2D correspondence but do not know the 3D motion
- We assume the current frame and the reference frame will not change dramatically

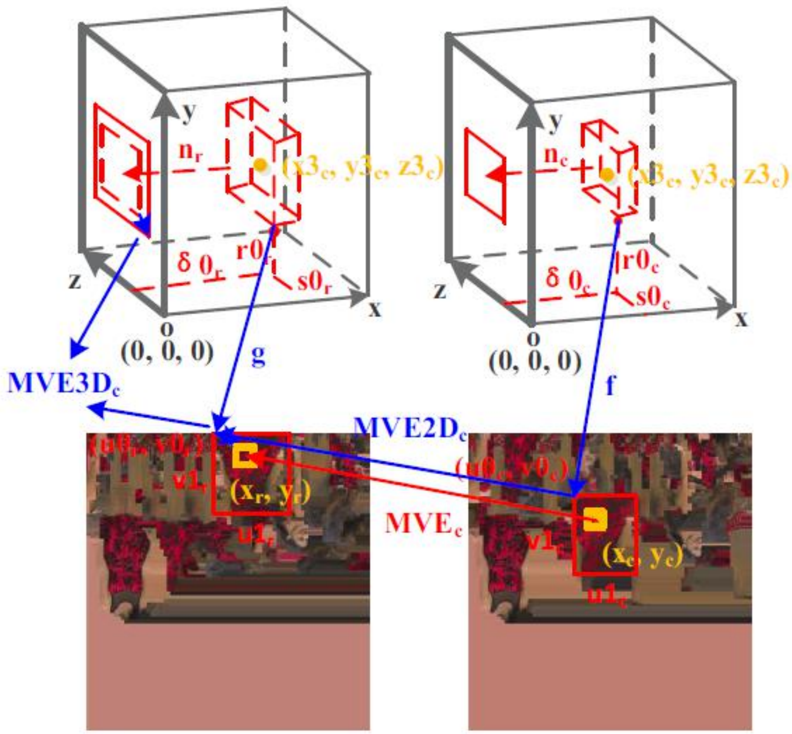
$$\boldsymbol{MVE_c} = \boldsymbol{g}(x3_c, y3_c, z3_c) - \boldsymbol{f}(x3_c, y3_c, z3_c)$$

- The problem is that (x3c,y3c,z3c) may not have a corresponding 2D point in the reference frame
  - We perform motion estimation which will increase the encoder and decoder complexity



Auxiliary information based motion prediction

- The previous method has the following two disadvantages
  - The high encoder and decoder complexity
  - It can only apply to the attribute



- The auxiliary information based motion prediction
  - The auxiliary information basically provides the coarse geometry
  - We use the 3D offset plus the 2D offset

Experiments setup

- The proposed algorithm is implemented in the V-PCC reference software and the corresponding HEVC reference software
- We test the all the dynamic point clouds defined in the common test condition including loot, redandblack, soldier, queen, longdress
- For the geometry, both point-to-point is point-to-plane are used
- For the attribute, the qualities of the luma, Cb, and Cr are considered

Experimental results on the overall scheme

- Overall scheme results

TABLE III
PERFORMANCE OF THE GEOMETRY-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR

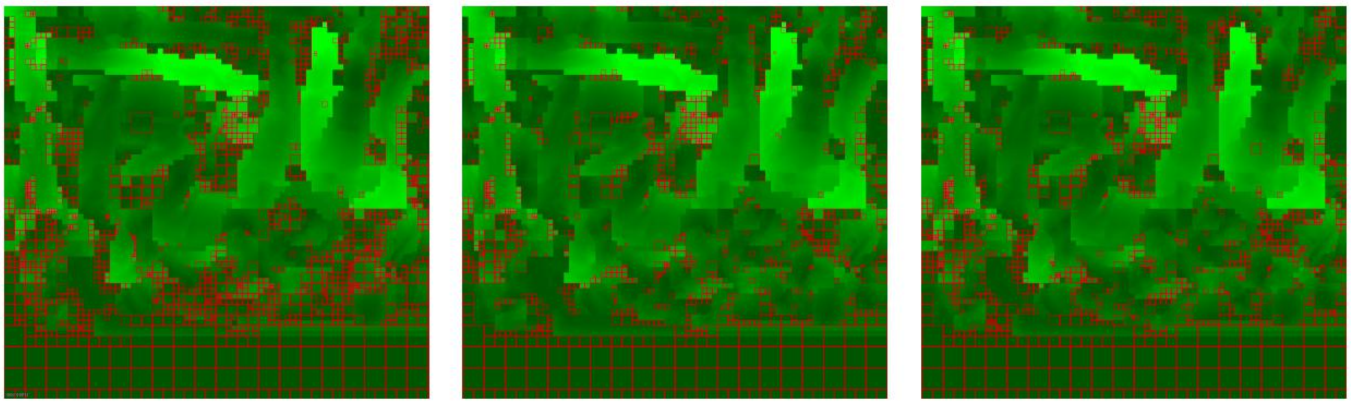
Test point cloud	Geom.BD-GeomRate D1	Geom.BD-GeomRate D2	Attr.BD-AttrRate Luma	Attr.BD-AttrRate Cb	Attr.BD-AttrRate Cr	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	0.0%	0.0%	–18.1%	–31.4%	–30.4%	–3.4%	–6.1%	–8.4%	–17.7%	–16.9%
RedAndBlack	0.0%	0.0%	–16.3%	–25.0%	–15.9%	–4.6%	–4.6%	–8.8%	–15.4%	–8.4%
Solider	0.0%	0.0%	–33.4%	–42.5%	–43.2%	–8.2%	–8.2%	–17.2%	–26.3%	–27.0%
Queen	0.0%	0.0%	–13.7%	–20.5%	–19.2%	–3.5%	–3.6%	–7.8%	–12.7%	–11.6%
LongDress	0.0%	0.0%	–9.8%	–13.5%	–12.3%	–3.7%	–3.7%	–6.4%	–9.5%	–8.4%
Avg.	0.0%	0.0%	–18.2%	–26.6%	–24.2%	–4.7%	–4.7%	–9.7%	–16.3%	–14.5%
Enc. time self	97%									
Dec. time self	98%									
Enc. time child	486%									
Dec. time child	337%									

TABLE IV
PERFORMANCE OF THE AUXILIARY-INFORMATION-BASED MOTION PREDICTION COMPARED WITH THE V-PCC ANCHOR UNDER THE NORMATIVE SOLUTION

Test point cloud	Geom.BD-GeomRate D1	Geom.BD-GeomRate D2	Attr.BD-AttrRate Luma	Attr.BD-AttrRate Cb	Attr.BD-AttrRate Cr	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	–4.0%	–3.9%	–16.3%	–26.4%	–28.5%	–6.3%	–6.2%	–9.6%	–16.7%	–17.9%
RedAndBlack	–1.0%	–1.1%	–12.2%	–18.9%	–10.9%	–4.0%	–4.1%	–7.2%	–12.1%	–6.2%
Solider	–8.0%	–7.9%	–31.3%	–41.4%	–40.4%	–13.6%	–13.4%	–19.8%	–28.7%	–28.1%
Queen	–5.9%	–5.9%	–11.8%	–17.0%	–15.7%	–7.3%	–7.3%	–9.1%	–12.9%	–11.8%
LongDress	–1.1%	–1.1%	–8.3%	–11.2%	–10.2%	–3.8%	–3.6%	–5.7%	–8.2%	–7.3%
Avg.	–4.0%	–4.0%	–16.0%	–23.0%	–21.1%	–7.0%	–6.9%	–10.3%	–15.7%	–14.3%
Enc. time self	100%									
Dec. time self	100%									
Enc. time child	98%									
Dec. time child	99%									

Performance Analysis

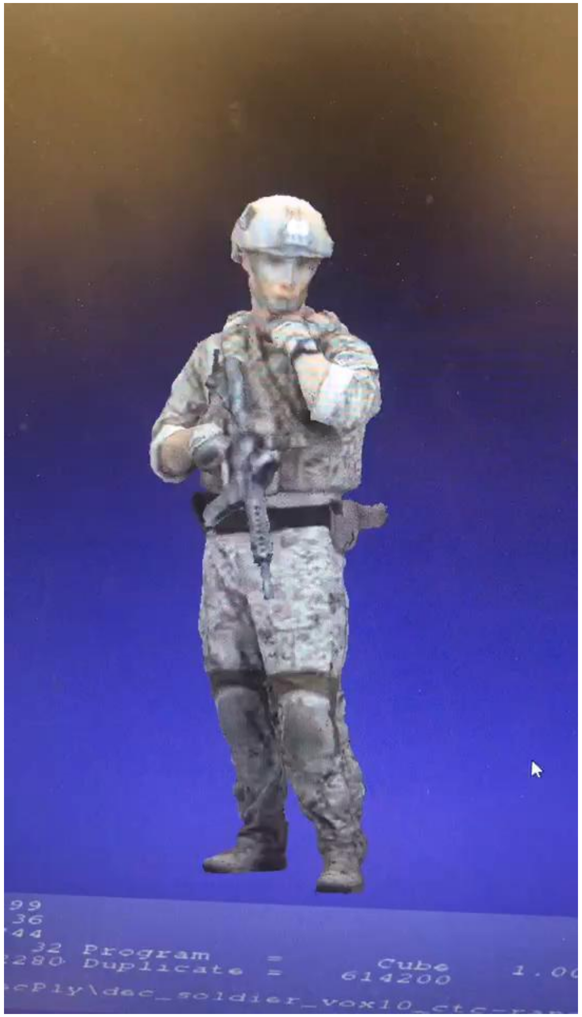
- Intra blocks reduce significantly, resulting in taking adv of inter coding efficiency



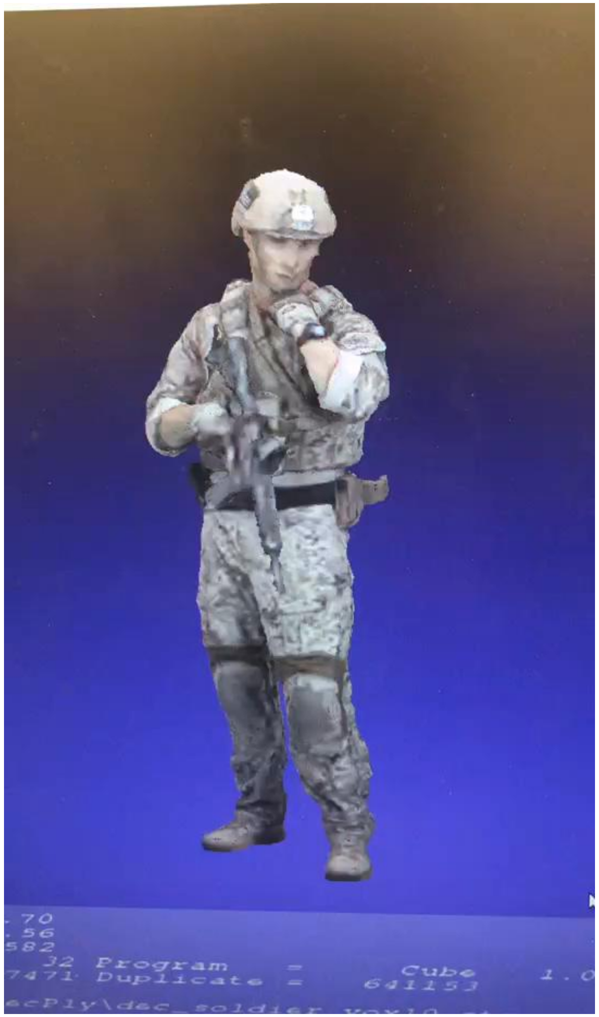
(a) Soldier Geometry Anchor (b) Soldier Geometry Normative (c) Soldier Geometry Non-normative

(d) Soldier Geometry Anchor (e) Soldier Geometry Normative (f) Soldier Geometry Non-normative

Subjective quality



Anchor



Proposed

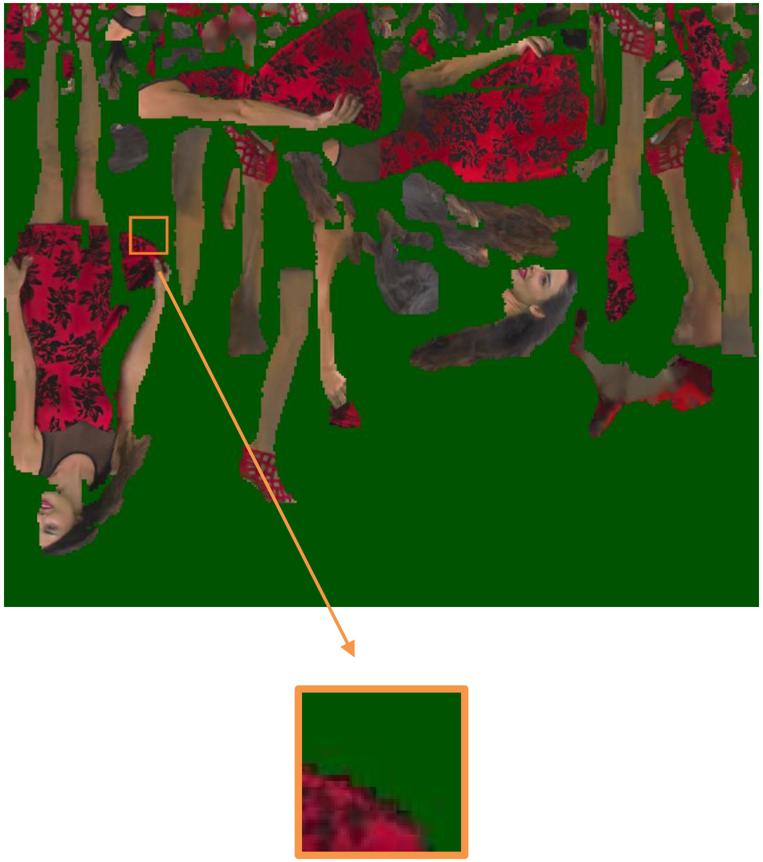
Occupancy Map Driven Rate-Distortion Optimization

- The current rate distortion optimization process in a video encoder such as HM is not handling the unoccupied pixels in a proper way

$$\min_{P} J = \sum_{i=1}^{N} D_i + \lambda R$$

⇩

For a block with both occupied and unoccupied pixels, all the pixels are treated as equal importance



Proposed occupancy-map-based RDO

- The unoccupied pixels are not beneficial for the reconstructed quality of the point cloud at all

- In the proposed solution, a distortion mask is added in the RDO to handle the unoccupied pixels

$$\min_{P} J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

where M_i is 1 when the current pixel is occupied, M_i is 0 when the current pixel is unoccupied

- This method is applied to intra/inter prediction and SAO

Intra prediction

- The RDO in intra prediction can be divided into three steps
  - INTRA Mode (Direction) Decision
    - The occupancy-map-based RDO is **not applied** as the residue bits are not counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} SATD_i + \lambda R_{dir}$$

  - Precise mode decision and residue Quadtree decision
    - The occupancy-map-based RDO is **applied** as the residue bits are counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

Inter prediction

- The inter mode can be divided into merge 2Nx2N and the other inter modes
  - Merge 2Nx2N/modes comparison
    - The occupancy-map-based RDO is **applied** as the residue bits are counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} D_i \times M_i + \lambda R$$

  - Other inter modes in Integer and fractional motion estimation processes or merge estimation
    - The occupancy-map-based RDO is **not applied** as the residue bits are not counted in the bit cost

$$\min_P J = \sum_{i=1}^{N} SAD_i / SATD_i + \lambda R_{motion}$$

Simulation setup

- We implement the proposed algorithm in V-PCC (TMC2-3.0) and the corresponding HEVC reference software to verify the performance of the proposed algorithm
- Follow the common test condition
  - Random access case and all intra case
- Test point cloud

Test point cloud	Frame rate	Number of points	Geometry precision	Attributes
Loot	30	~780000	10bit	RGB
RedAndBlack	30	~700000	10bit	RGB
Soldier	30	~1500000	10bit	RGB
Queen	50	~1000000	10bit	RGB
Longdress	30	~800000	10bit	RGB

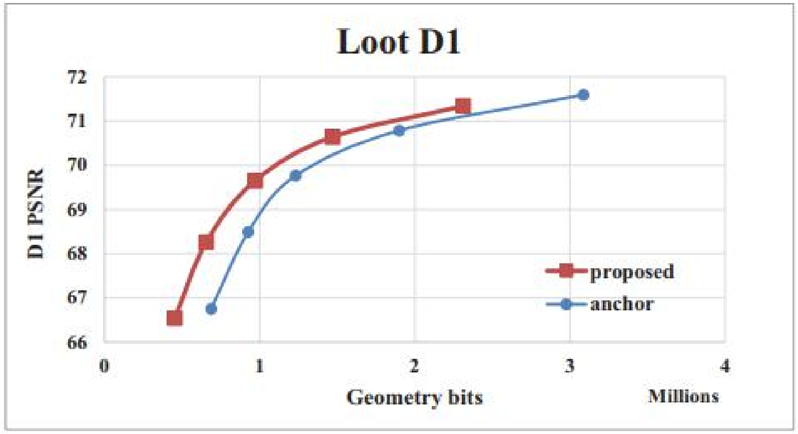
Experimental results

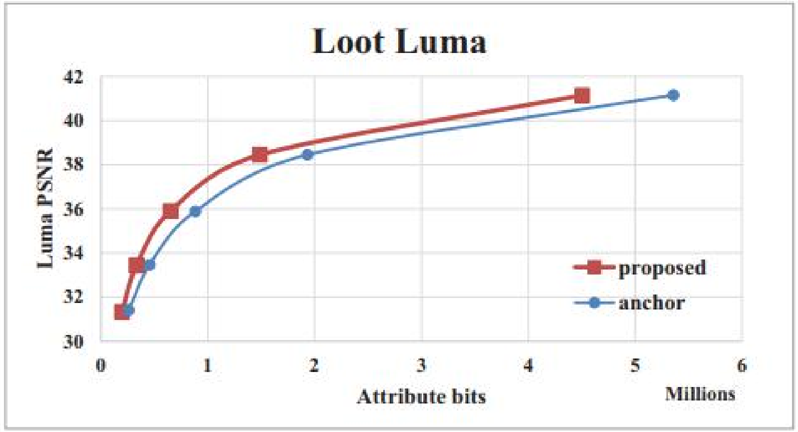
- Random access case

Test point cloud	Geom.BD-Rate		Attr.BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	-16.3%	-16.4%	-24.3%	-18.2%	-19.3%
RedAndBlack	-6.6%	-7.2%	-12.2%	-9.8%	-12.3%
Soldier	-15.8%	-16.0%	-16.8%	-9.4%	-9.0%
Queen	-13.4%	-13.2%	-15.7%	-11.2%	-10.5%
Longdress	-7.5%	-7.8%	-7.9%	-7.7%	-7.2%
Avg.	-11.9%	-12.1%	-15.4%	-11.3%	-11.7%
Enc. self	101%				
Dec. self	99%				
Enc. child	88%				
Dec. child	88%				

Experimental results

- Examples of R-D curves in random access case



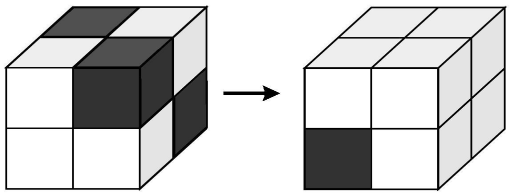


Outline

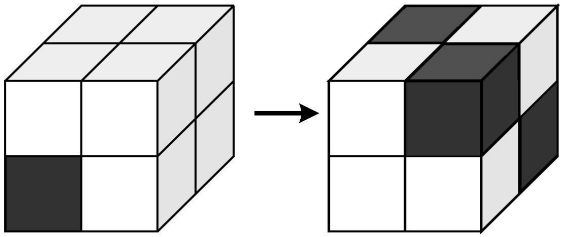
- Short Self Intro
- Research Motivation and Highlights
- Scalable Point Cloud Geometry Compression
- Video Projection Based Point Cloud Compression
- Post-Processing: Point Cloud Scaling
- Summary

Point Cloud Scaling

- Scaling by voxelization



Voxel merging due to quantization



Upsample using occupancy prediction

10-bit

9-bit

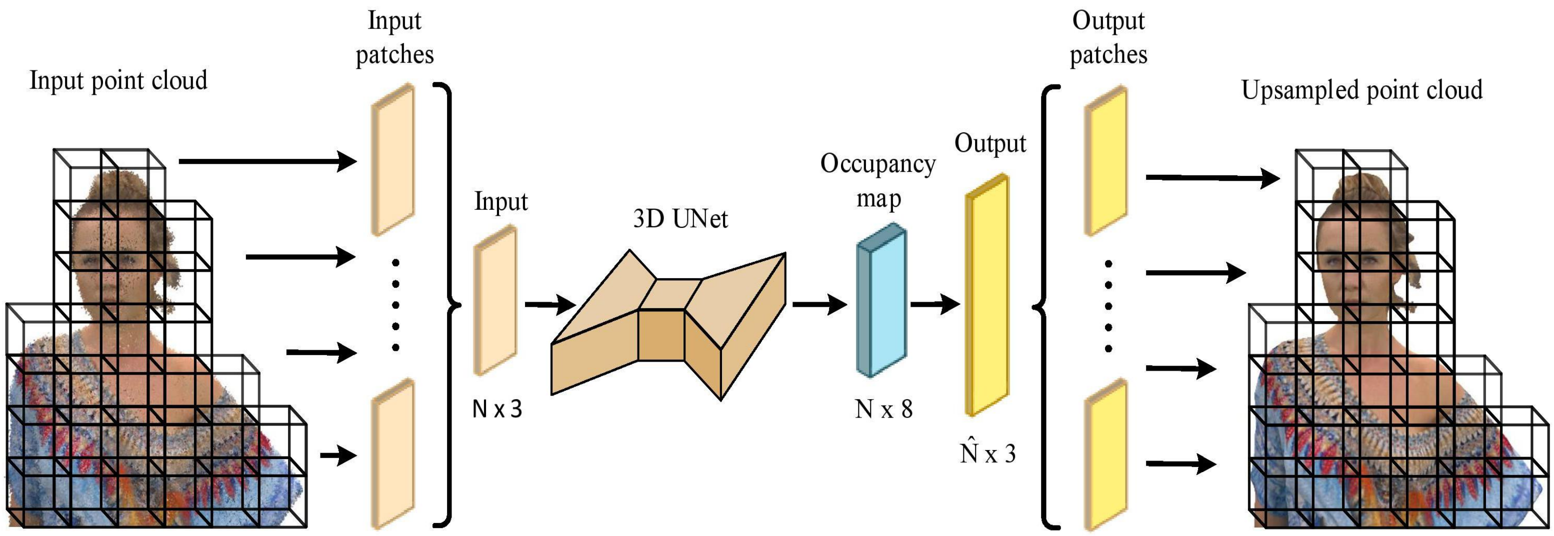
8-bit

Upscaling a point cloud

- Current point cloud lossy compression and processing schemes suffer from quantization loss which results in a coarser sub-sampled representation of point cloud.
- We solve the problem of points lost during voxelization by performing geometry prediction across spatial scale using deep learning architecture.
- Helpful tool for point cloud compression as well as display adaptation.

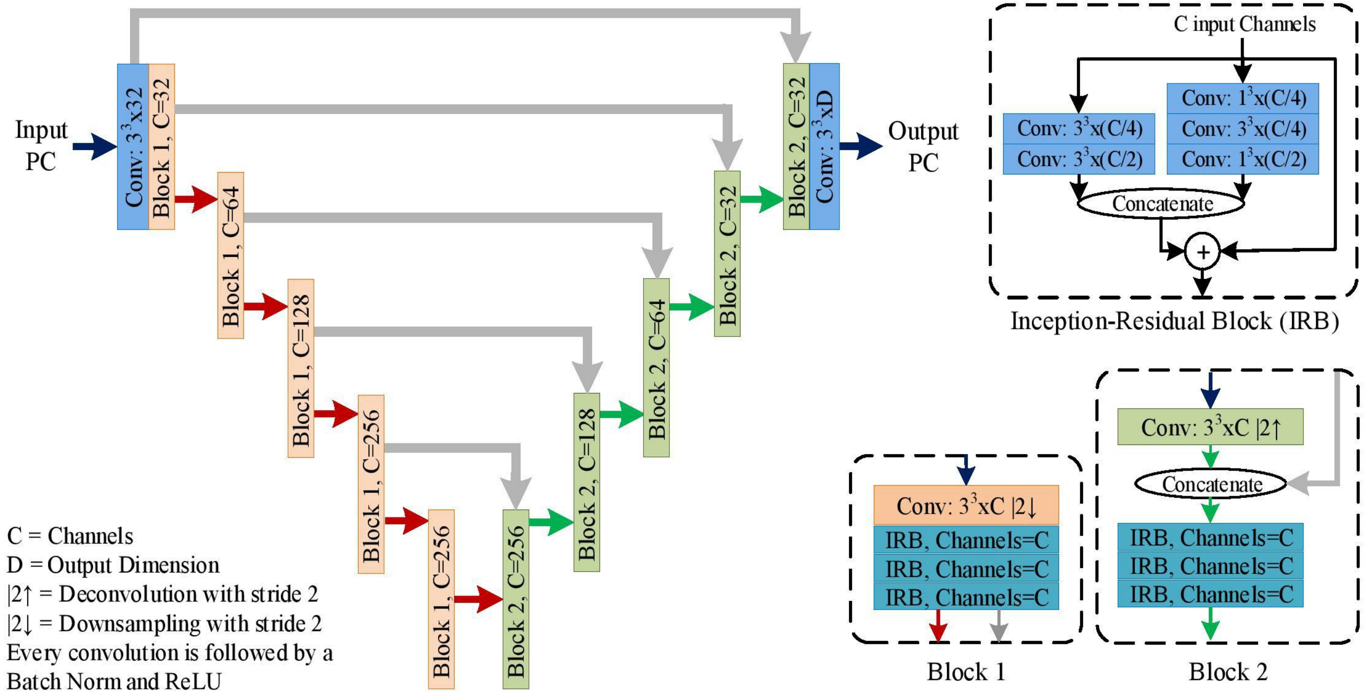
Deep Learning Solution

- Using a SparseNet backbone:
  - patch based feature embedding with 3D Unet
  - loss function as a voxel octree partition code prediction loss - 8bit binary cross entropy
  - different from typical chanfer distance and earth moving distance loss based and better captures the problem.



3D Unet Design Details

- Including our inception residual block units
  - PointCNN/PointNet type backbone not able to handle such large data set



Significant Performance Gains over SOTA

- Scaling Performance:

TABLE I: Average PSNR (dB) results.

qs	Input Quantized PC	Output PC	Difference	Increase in points
4/3	64.6646	73.8630	9.1984	x1.7
2	63.2080	72.0758	8.8678	x3.7
4	58.0077	65.1890	7.1813	x13.8

- Compared with current SOTA like PU-Net and similar ones, we have 4~5dB advantage

Subjective Results

(a) (b) (c)

Fig. 5: (a) Original point cloud, (b) Quantized point cloud with qs = 2, (c) Upsampled point cloud.

(a) (b) (c)

Fig. 6: (a) Original point cloud, (b) Quantized point cloud with qs = 4, (c) Upsampled point cloud.

Summary

- Immersive Visual Communication can enable many crucial applications in remote medicine, education, and military use cases.
- For 3D sensing/Auto-driving, geometry is the key, BTQT is a good framework with room for entropy coding optimization (LSTM), and RDO
- vPCC deals with immersive content, current MPEG vPCC has many in-efficiency, we introduced advanced motion model, occupancy map based RDO to significantly improve the over all performance
- Deep learning point cloud geometry super resolution, learn a large scale feature embedding with a novel octree node occupany loss function
- Plenoptic (multi-attributes) point cloud coding with light-field like coding scheme, inter-view prediction yields very good results, for even higher quality immersive experience